

GENERATING HYPERLINKS AND ANCHOR TEXT IN HTML AND NON-HTML DOCUMENTS

ABSTRACT OF THE DISCLOSURE

Systems and methods for generation of hyperlinks and anchor text from data such as reference text in HTML and in non-HTML documents are disclosed. The method generally includes locating a text reference in a source document, searching using a search engine for a target document relating to the text reference, computing anchor text from the text reference, generating a hyperlink to the target document, and associating the hyperlink with the computed anchor text. The locating and/or computing may be based on a respective statistical model of text formatting and/or lexical cues. The text reference may be parsed into pieces such that the searching, computing, generating, and associating are performed for each piece of text. The source document may be an HTML or non-HTML document. The text reference may be a reference to, for example, a paper, article, company, institution, product, search engine, image, object, and geographical location.